

2021 IEEE 6th International Conference on Big Data Analytics (ICBDA 2021)

**Xiamen, China
5 – 8 March 2021**



**IEEE Catalog Number: CFP21TPE-POD
ISBN: 978-1-6654-3037-1**

**Copyright © 2021 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

| | |
|-------------------------|-------------------|
| IEEE Catalog Number: | CFP21TPE-POD |
| ISBN (Print-On-Demand): | 978-1-6654-3037-1 |
| ISBN (Online): | 978-1-6654-1513-2 |

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

Table of Contents

2021 IEEE the 6th International Conference on Big Data Analytics (ICBDA 2021)

| | |
|----------------------------|------|
| Preface..... | viii |
| Conference Committees..... | ix |

Big Data Science and Key Technologies

| | |
|--|----|
| Big Data Driven Model for New York Taxi Trips Analysis | 1 |
| <i>Xuqiao Yu</i> | |
| Research on the Core Technology of Education Big Data Based on Data Mining | 5 |
| <i>Chun-li Wang</i> | |
| Research and Design of Credit Risk Assessment System Based on Big Data and Machine Learning | 9 |
| <i>Song Wen, Bi Zeng, Wenxiong Liao, Pengfei Wei, Zhihao Pan</i> | |
| High Dimensional Complexity of Dynamical System in the Big Data | 14 |
| <i>Li Zhang, Jianquan Sun, Shutang Liu</i> | |
| Research on Education Big Data Security Strategy under Network Environment..... | 19 |
| <i>Tian Fupeng</i> | |
| Research and Analysis of Student Portrait Based on Campus Big Data | 23 |
| <i>Xiaoying Li, Shouwu He</i> | |
| Research on Agricultural Products Intelligent Recommendation Based on E-commerce Big Data | 28 |
| <i>Wu Junhui, Shen Chushan, Wu Yusheng, Chen Jie, Lin Kaiyan, Si Huiping</i> | |
| An Improved Error-Based Pruning Algorithm of Decision Trees on Large Data Sets | 33 |
| <i>Yi Peng, Yu-Tong Lu, Zhi-Guang Chen</i> | |
| Doctor Recommendation Model for Pre-Diagnosis Online in China: Integrating Ontology Characteristics and Disease Text Mining..... | 38 |
| <i>Chunhua Ju, Shuangzhu Zhang</i> | |

Machine Learning

| | |
|--|----|
| The Use of Features to Enhance the Capability of Deep Reinforcement Learning for Investment Portfolio Management | 44 |
| <i>Xiaotian Ren, Zhengyong Jiang, Jionglong Su</i> | |
| Property Prediction of Medical Magnesium Alloy based on Machine Learning | 51 |
| <i>Na Li, Siyue Zhao, Zhigang Zhang</i> | |
| Using Artificial Intelligence to Predict Legislative Votes in the United States Congress..... | 56 |
| <i>Anasse Bari, William Brower, Christopher Davidson</i> | |
| Stock Price Prediction with ARIMA and Deep Learning Models | 61 |
| <i>Zihao Gao</i> | |
| Application of Machine Learning Model and Hybrid Model in Retail Sales Forecast | 69 |
| <i>Haichen Jiang, Jiatong Ruan, Jianmin Sun</i> | |
| A Deep Residual Shrinkage Neural Network-based Deep Reinforcement Learning Strategy in Financial Portfolio Management | 76 |
| <i>Ruoyu Sun, Zhengyong Jiang, Jionglong Su</i> | |
| Prediction of the Electronic Work Function by Regression Algorithm in Machine Learning | 87 |
| <i>Na Li, Tianxin Zong, Zhigang Zhang</i> | |
| Application of Features and Neural Network to Enhance the Performance of Deep Reinforcement Learning in Portfolio Management | 92 |
| <i>Fengchen Gu, Zhengyong Jiang, Jionglong Su</i> | |

Data Analysis and Data Management

| | |
|---|-----|
| An Interpretability Method of Workflow Power Consumption Based on Feature Contribution | 98 |
| <i>Wanpeng Ji, Xusheng Zhang, Ziyu Shen, Bin Xia, Binghui Liu, Huizhong Shan, Shajia Shan, Zheng Liu</i> | |
| A Champion Model of Men’s 100m Sprint Based on Functional Data Analysis..... | 103 |
| <i>Dandan Cui, Zili Jiang</i> | |
| Fast Object Detection for Remote Sensing Data Using a Mobile Device | 108 |
| <i>Jian Huang, Xiushan Zhang, Hao Xu</i> | |
| Research and Application of Voltage Transformer Error Prediction Technology on Multi-Time Scale Data Analysis Method..... | 113 |
| <i>Zeng Feitong, Tong Guanghua</i> | |
| Anomaly Detection for Time Series Data Stream | 118 |
| <i>Qifan Wang, Bo Yan, Hongyi Su, Hong Zheng</i> | |
| Microarray Data Classification Based on Neighbourhood Components Analysis Projection Method | 123 |
| <i>Chuanlei Zhang, Lixin Liu, Shanwen Zhang, Chengliang Huang</i> | |

Entropy-based Traffic Congestion Propagation Pattern Mining with GPS Data 128
Xinyuan Sui, Yaying Zhang

Comparison and Evaluation of the Combinations of Feature Selection and Classifier on Microarray Data 133
Chaokun Yan, Jun Zhang, Xi Kang, Zhengze Gong, Jianlin Wang, Ge Zhang

Blockchain and Data Security

Study on Library Individualized Information Security under the Background of Big Data 138
Yang Fangjing

Blockchain-Based Distributed Data Integrity Auditing Scheme 143
Han Baofu, Li Hui, Wei Chuansi

Research and Implementation of Embedded Database Encryption System based on Lightweight Cryptographic Algorithms 150
Jishun Liu, Yan Zhang, Zuwei Yu, Xiye Gu

Ocean Data Sharing Based on Blockchain 155
Xiwen Cheng, Fengzhong Qu

Towards Smart Contracts for Verifying DNN Model Generation Process with the Blockchain 160
Hirotsugu Seike, Yasukazu Aoki, Noboru Koshizuka

Text Analysis and Clustering Algorithm

Analysis and Research on Enterprise Resumption of Work and Production Based on K-means Clustering 169
Wang Peiyi, Ma Longfei, Li Xianglong, Zhang Lu, Qin Hao, Yao Bin, Ding Yifeng

Point of Interest Recommendation Acceleration Using Clustering 175
Huida Jiao, Fan Mo, Hayato Yamana

Trajectory Similarity Measure Design for Ship Trajectory Clustering 181
Yuan-qiang Zhang, Guo-you Shi

Time Series Classification via Enhanced Temporal Representation Learning 188
Kun Wang, Chun Wang, Yunxiao Wang, Wei Luo, Peng Zhan, Yupeng Hu, and Xueqing Li

Improving Text Classification Using Knowledge in Labels 193
Cheng Zhang, Hayato Yamana

Incremental Real-time Learning Framework for Sentiment Classification: Indian General Election 2019, A Case Study 198
Sharmistha Chatterjee, Sushmita Gupta

Modular Policy Evaluation System: A Policy Evaluation Framework Based on Text Mining 204
Yang Gao, Yang yang Li, Yaojun Wang

High-Performance Computing and Algorithms

Research on the Learning Behavior of Students in Blended Learning Mode 210
Zhu Long, Xingsen Mu, Chunhong Song, Dan Tian

| | |
|--|-----|
| Prediction Algorithm for Drugstore Consumption Members | 214 |
| <i>Yuan Chang, Tao Li</i> | |
| Improved K-means Based on Density Parameters and Normalized Distance | 219 |
| <i>Xing Che, Hengyi Tao, Zihan Shi</i> | |
| GSP Distributed Deep Learning Used for the Monitoring System | 224 |
| <i>Zhongming Pan, Yigui Luo, Wei Sha, Yin Xie</i> | |
| A Deep Deterministic Policy Gradient-based Strategy for Stocks Portfolio Management | 230 |
| <i>Huanming Zhang, Zhengyong Jiang, Jionglong Su</i> | |
| A Novel Disease Prediction Method Based on Inductive Representation Learning | 239 |
| <i>Jianlin Wang, Xu Ma, Renyu Song, Chaokun Yan, Huimin Luo</i> | |
| Innovative Teaching Models of Higher Mathematics in Smart Classroom..... | 244 |
| <i>Jie Shen, Qinglin Wu</i> | |
| A Game-Incorporated Alternative Least Squares-Based Approach to Latent Factor Analysis..... | 249 |
| <i>Rui Che, Ye Yuan</i> | |
| Computer Theory and Information Engineering | |
| Framework for Analyzing and Visualizing Topics and Sentiments on Social Media: The Case of MH17 Tweets | 257 |
| <i>Jouko Vankka, Alexandr Vesselkov, Heikki Myllykoski, Onni Kosomaa</i> | |
| STAR: A Zone Translation Scheme to Improve the Performance of Host-Aware SMR..... | 267 |
| <i>Wenguo Liu, Hui Li, Lingfang Zeng</i> | |
| GPU Accelerated Matrix Factorization for Recommender Systems | 272 |
| <i>Doruk Kilitcioglu, Nicholas Greenquist, Mohamed Zahran, Anasse Bari</i> | |
| Collaborative Replication for Mobile Short Video Sharing Services..... | 282 |
| <i>Yuhan Hu, Lan Ding, Ye Tian</i> | |
| Multi-label Image Recognition with Asymmetric Cooccurrence Dependency Graphs | 287 |
| <i>Yuhang Qi, Yuchun Guo, Yishuai Chen</i> | |
| Fast Range Query of Fault Equipment Location Based on Breadth and Depth Ratio of the Binary Tree Model.... | 295 |
| <i>Yuying Zhang, Chao-Hsien Hsieh, Qing Zhang, Xiaoyu Wu, Tong Ji</i> | |
| Research on the Demand Distribution of Aviation Equipment Maintenance Spare Parts | 301 |
| <i>Guo Feng, Wang Jingjing, Wang Qingbin, Zhang Suqin</i> | |
| Social Construction of Work Rhythm for Digital Transformation..... | 305 |
| <i>Jiaona Xiang, Shuxian Chen, Lu Cheng, Zongqiang Ren</i> | |
| A Novel Perspective to Mining Online Hotel Reviews Based on Biterm Topic Model | 310 |
| <i>Qianqian Ma, Huiying Du, Zhiyuan Wang</i> | |

Information Network and System

| | |
|---|-----|
| Named Entity Recognition Using BERT with Whole World Masking in Cybersecurity Domain | 316 |
| <i>Shicheng Zhou, Jingju Liu, Xiaofeng Zhong, Wendian Zhao</i> | |
| Network Fault Classification Techniques Based on Transfer Learning | 321 |
| <i>Yang Shan, Zhou Zheng, Liu Zheng, Jiang Yan, He Liang, Liu Yunpeng</i> | |
| The Edge Computing Cloud Architecture Based on 5G Network for Industrial Vision Detection | 328 |
| <i>Tielin Lu, Zitian Fan, Yue Lei, Yujia Shang, Chunxi Wang</i> | |
| Malware Variant Detection Based on Decomposed Deep Convolutional Network..... | 333 |
| <i>Jianbin Mai, Chunjie Cao, Fangfei Shi, Xiaoqing Chen</i> | |
| Deep Sentiment Learning Network for Temporalaware Recommendation Based on User Reviews | 339 |
| <i>Xinxin Li, Tianqi Shang, Dezhong Peng, Xiaoyu Shi</i> | |
| Web User Behavior Trust Evaluation Model Based on Fuzzy Petri Net..... | 344 |
| <i>Zenan Wu, Liqin Tian, Zhigang Wang, Yan Wang</i> | |
| Open Source Galaxy: Heterogeneous Information Networks in Social Coding | 349 |
| <i>Tianyi Zhou, Wei Wang, Shengyu Zhao</i> | |
| Author Index..... | 356 |